



Europäisches
Patentamt

European
Patent Office

Office européen
des brevets

REC'D 09 JAN 2004

WIPO

PCT

Bescheinigung

Certificate

Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

03100016.9

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

**PRIORITY
DOCUMENT**

SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)



Anmeldung Nr:
Application no.: 03100016.9
Demande no:

Anmeldetag:
Date of filing: 08.01.03
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Philips Intellectual Property & Standards
GmbH
Steindamm 94
20099 Hamburg
ALLEMAGNE
Koninklijke Philips Electronics N.V.
Groenewoudseweg 1
5621 BA Eindhoven
PAYS-BAS

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

Vorrichtung und Verfahren zum Unterdrücken von Echo, insbesondere bei Telefonen

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s) revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

**Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:**

H04M9/08

Am Anmeldestag benannte Vertragstaaten/Contracting states designated at date of
signature/Etats contractants désignés lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LU MC NL PT SE SI SK TR LI

BESCHREIBUNG

Vorrichtung und Verfahren zum Unterdrücken von Echo, insbesondere bei Telefonen

Die vorliegende Erfindung betrifft eine Vorrichtung und ein entsprechendes Verfahren zum Unterdrücken von Echo für eine Anordnung zum Übertragen von Audiosignalen,

5 insbesondere gesprochener Sprache, wie beispielsweise Telefonen, mit:

- einer Echoreduktionseinheit, die zwischen einem Eingangskanal zum Empfangen eines von einem fernen Ende kommenden Eingangs-Audiosignals und einem Ausgangskanal zum Ausgeben eines Ausgangs-Audiosignals angeordnet ist, zur Unterdrückung eines in dem Ausgangs-Audiosignal enthaltenen Echosignals,
- 10 - einer Sprachaktivitätsdetektionseinheit zur Detektion eines in dem Eingangs-Audiosignal enthaltenen Sprachsignals, und
- einer Steuereinheit zur Einstellung eines Echounterdrückungsfaktors der Echo-Unterdrückungseinheit für die Unterdrückung von Echo.

15 Jedes Telefon hat einen Lautsprecher und ein Mikrofon, die nicht vollständig voneinander akustisch entkoppelt sind, wodurch Echos entstehen können, die der ferne Sprecher am fernen Ende hört. Dieses Problem tritt insbesondere bei Freisprecheinrichtungen, aber auch bei Mobilfunktelefonen und schnurlosen Telefonen aufgrund der geringen Gehäusegröße und des relativ großen Abstandes zwischen dem Mund des Sprechers und 20 dem Mikrofon auf. Um dieses sogenannte akustische Echo zu vermeiden, werden unterschiedliche Ansätze oder Kombinationen davon benutzt. Mittels eines Halb-Duplex-Schalters, der denjenigen Sprecher dämpft, der mit leiserer Stimme spricht, können zwar Echos zuverlässig unterdrückt werden, es besteht jedoch nicht die Möglichkeit, dass gleichzeitig in beiden Richtungen gesprochen wird (sogenanntes Voll-Duplex).

25 Mittels einer Echolöscheinheit, die eine Schätzung des Echosignals von dem Mikrofonsignal subtrahiert, ist zwar eine Voll-Duplex-Kommunikation möglich, Echos werden damit jedoch nicht immer vollständig unterdrückt. Eine weitere Möglichkeit ist ein Echounterdrücker, der das Ausgangssignal in dem Mikrofonpfad nur dann dämpft, wenn Echos unterdrückt werden müssen.

Aus der WO 99/35812 A1 sind ein Verfahren und eine Vorrichtung zur Steuerung der Echounterdrückung in einem Kommunikationssystem bekannt. In einer ersten Ausgestaltung ist der akustische Echounterdrücker deaktiviert, wenn ein Sprachaktivitätsdetektor anzeigt, dass das von dem fernen Sprecher kommende Audiosignal keine

5 Sprachkomponente enthält. In einer weiteren Ausgestaltung wird die Echounterdrückung nicht vollständig deaktiviert, sondern auf ein niedriges Unterdrückungsniveau eingestellt, welches ausreichend ist, um Nachhall zu unterdrücken.

Bei dem bekannten System wird also die Unterdrückung sofort zurückgeschaltet, wenn

10 das von dem fernen Sprecher kommende Eingangs-Audiosignal keine Sprachsignal enthält. Anfangs können dabei aber noch Echos „unterwegs“ sein wegen des Nachhalls. Der ferne Sprecher kann also ein kurzes Echo hören, nachdem er selbst aufhört zu sprechen oder auch nach sogenannten Plosivlauten, vor allem nach den Lauten „t“, „k“ und „z“. Darüber hinaus kann sich insbesondere beim Start des Systems, also bei

15 Beginn eines Telefongesprächs, ein weiteres Problem ergeben, wenn, was häufig der Fall ist, die Echoreduktionseinheit (echo reduction unit) eine sogenannte Echokompensationseinheit (echo cancelling unit) mit adaptivem FIR-Filter (finite impulse response filter) aufweist. Bei Beginn des Telefongesprächs ist ein solches Filter nämlich noch nicht adaptiert, und es wird keine Echokompensation erzielt. Die

20 Echoreduktion eines solchen FIR-Filters steigt erst langsam an und kann nur bei ferner Sprache adaptieren. Dies hat zur Folge, dass am Anfang noch Echos durchkommen und dass erst nach einer Zeit von einer halben bis einer Sekunde die Echounterdrückung (echo suppression) vollständig wirkt.

25 Der vorliegenden Erfindung liegt deshalb die Aufgabe zugrunde, eine Vorrichtung und ein entsprechendes Verfahren zur wirksamen Unterdrückung von Echos bei einer Audioübertragungsvorrichtung, insbesondere einem Telefon, vorzusehen, wobei die oben beschriebenen Probleme vermieden werden sollen, bei dem also insbesondere dann, wenn das von einem fernen Ende kommende Eingangs-Audiosignal kein

30 Sprachsignal enthält, wirksam und schnell Echos unterdrückt werden.

Diese Aufgabe wird erfindungsgemäß durch eine Vorrichtung gemäß Anspruch 1 gelöst, die dadurch gekennzeichnet ist, dass die Steuereinheit derart ausgebildet ist, dass der Echounterdrückungsfaktor von einem während des Vorliegens eines Sprachsignals 5 in dem Eingangs-Audiosignal eingestellten hohen Echounterdrückungswert schrittweise und kontinuierlich auf einen niedrigen Echounterdrückungswert reduziert wird, wenn von der Sprachaktivitätsdetektionseinheit detektiert wird, dass das Eingangs-Audiosignal kein Sprachsignal enthält.

10 Ein entsprechendes Verfahren zur Echo reduktion ist in Anspruch 11 angegeben. Ein Computerprogramm zur Implementierung des Verfahrens ist in Anspruch 12 angeben. Vorteilhafte Ausgestaltungen der Erfindung sind in den Unteransprüchen angegeben.

Der Erfindung liegt die Erkenntnis zugrunde, den Echounterdrückungsfaktor nicht 15 abrupt vollständig auf Null oder einen niedrigen Wert herunterzuschalten, wenn das Eingangs-Audiosignal kein Sprachsignal enthält, sondern stufenweise und kontinuierlich, da ansonsten Echos aufgrund von Nachhall übertragen werden, wenn der ferne Sprecher aufhört zu sprechen oder insbesondere auch nach sogenannten Plosivlauten. Derart störende Echos werden bei der erfindungsgemäßen Vorrichtung 20 vermieden, da der Echounterdrückungsfaktor in dem Moment, wo der ferne Sprecher aufhört zu sprechen, oder nach einem Plosivlaut, nachdem insbesondere an Wortenden eine kurze Pause kommt, erst allmählich abgesenkt wird und nicht abrupt wie bei der bekannten Vorrichtung. Zu diesem Zeitpunkt werden Echos also noch ausreichend stark unterdrückt, um bei dem fernen Sprecher keinen Nachhall ankommen zu lassen. Dies 25 führt zu einer wesentlich angenehmeren Sprachübertragung bei der erfindungsgemäßen Vorrichtung.

In einer bevorzugten Ausgestaltung ist vorgesehen, das Abklingen von dem hohen Echounterdrückungswert auf den niedrigen Echounterdrückungswert mittels einer 30 exponentiellen Funktion zu steuern. Dies bildet in etwa den wahren Zeitverlauf der Echoenergie nach, da diese bekanntlich auch exponentiell mit der Zeit abnimmt, und

führt somit zu einer optimalen Echounterdrückung in den genannten Zeitpunkten.

Eine bevorzugte Funktion für die Steuerung des Echounterdrückungsfaktors in den genannten Zeitpunkten ist in Anspruch 3 angegeben.

5 In einer weiteren Ausgestaltung ist vorgesehen, das Zurückschalten des Echounterdrückungsfaktors um eine bestimmte Totzeit zu verzögern, also noch während dieser Totzeit den hohen Echounterdrückungswert eingestellt zu lassen und erst danach den Echounterdrückungsfaktor abzusenken. Auch dadurch kann Nachhall wirksam

10 unterdrückt werden. Allerdings ist eine derart hohe Echounterdrückung während der Totzeit nicht zwingend erforderlich und trägt nur zu einer unnötigen Verschlechterung der Übertragungsqualität des von dem nahen Ende (bzw. dem nahen Sprecher) kommenden Audiosignals, insbesondere eines Sprachsignals, bei.

15 In einer Weiterbildung ist eine weitere Sprachaktivitätsdetektionseinheit vorgesehen zur Detektion, ob das von dem nahen Ende kommende Audiosignal ein Sprachsignal des nahen Sprechers enthält. In diesem Fall ist die Steuereinheit so ausgebildet, dass sie bei Vorliegen von Sprachsignalen sowohl in dem Eingangs-Audiosignal als auch in dem von dem nahen Ende kommenden Audiosignal einen mittleren Echounterdrückungswert

20 einstellt, während nur bei Nichtvorhandensein eines Sprachsignals in dem Eingangs-Audiosignal der maximale Echounterdrückungswert eingestellt wird. Im ersten Falle werden zwar Echos nicht vollständig unterdrückt; das von dem nahen Ende kommende Sprachsignal wird aber doch hörbar, gegebenenfalls verzerrt, übertragen.

25 Bevorzugt wird bei der erfindungsgemäßen Vorrichtung eine Echoreduktionseinheit verwendet, die ein adaptives FIR-Filter aufweist zur Erzeugung einer Schätzung des in dem von dem nahen Ende kommenden Audiosignal enthaltenen Echos des Eingangs-Audiosignals. Ein solches FIR-Filter wird häufig auch in einer Echokompensationseinheit verwendet, die das von dem nahen Ende kommende Signal

30 nicht verändert. Die genannte Schätzung wird im Allgemeinen von dem von dem nahen

Ende kommende Audiosignal subtrahiert, bevor das von dem nahen Ende kommende Audiosignal, an das ferne Ende übertragen wird. Die Leistung dieser Echoschätzung dient zusätzlich auch als Eingangssignal für eine dem FIR-Filter nachgeschaltete Echounterdrückungseinheit. Insbesondere beim Start der Kommunikation, also beim

5 Beginn eines Telefongesprächs, ist ein solches adaptives FIR-Filter jedoch noch nicht adaptiert und die Echounterdrückungseinheit wird mit einer zu geringen Schätzung der Echoleistung gespeist, was eine zu geringe Echounterdrückung zur Folge hat. Erfindungsgemäß ist deshalb in einer vorteilhaften Ausgestaltung vorgesehen, dass die Koeffizienten des FIR-Filters beim Start einer Kommunikation ungleich Null

10 initialisiert werden, um eine realistische Leistung im Schätzsignal zu erreichen, wodurch eine ausreichende Echounterdrückung auch in der Startphase erreicht wird.

Eine weitere, davon ausgehende Ausgestaltung sieht vor, in der Startphase den Echounterdrückungsfaktor entsprechend dem anfänglich schlechten Adoptionszustand

15 des FIR-Filters zu erhöhen, um die zu geringe Echokompensation in der Startphase auszugleichen. Erst allmählich wird dann, beispielsweise mittels einer Zeitsteuerung, der Echounterdrückungsfaktor auf einen statischen hohen Echounterdrückungswert abgesenkt, der dann beibehalten bleibt, solange in dem Eingangs-Audiosignal ein Sprachsignal enthalten ist. Beispielhaft kann die Zeitsteuerung so ausgestaltet sein, dass

20 die Absenkung auf den statischen hohen Echounterdrückungswert nach etwa 0,5 bis 1 Sekunde, vorzugsweise nach etwa 0,8 Sekunden, erreicht ist. Dadurch werden in der Startphase, also am Anfang eines Telefongesprächs, Echos mit absoluter Sicherheit vermieden. Während dieser Zeit ist die Vollduplex-Fähigkeit, also die Fähigkeit, dass der ferne Sprecher hört, wenn ihn der nahe Sprecher unterbrechen will oder eine kurze

25 Bemerkung macht, während der ferne Sprecher aktiv ist, vom nahen zum fernen Sprecher hin zwar eingeschränkt. Dies ist im Allgemeinen jedoch kein Nachteil, da in den ersten Sekunden eines Telefonats selten Gegensprechen stattfindet, da üblicherweise jeder Teilnehmer zunächst seinen Namen oder einen Gruß übermittelt.

30 Beispielhafte Werte für die genannten Echounterdrückungswerte sind in weiteren Unteransprüchen angegeben.

Die Erfindung wird nachfolgend anhand der Zeichnungen näher erläutert. Es zeigen:

Fig. 1 ein Blockschaltbild einer erfindungsgemäßen Echo reduktionsvorrichtung,

Fig. 2 Zeitverläufe des Echo unterdrückungsfaktors und der erzielten

5 Echo unterdrückung bei der erfindungsgemäßen Vorrichtung und

Fig. 3 ein Blockschaltbild einer erfindungsgemäßen verwendeten
Sprachaktivitätsdetektionseinheit.

Fig. 1 zeigt als Blockschaltbild eine erfindungsgemäße Echo reduktionsvorrichtung.

10 Diese ist in der gezeigten Ausführung bei einer Anordnung zur Übertragung von Audiosignalen, also beispielsweise bei einem Mobilfunktelefon, einer Freisprecheinrichtung, einem Funkgerät oder einem drahtlosen Telefon, eingesetzt. Dabei wird ein von einem fernen Ende (fernen Sprecher) auf einem Eingangskanal 1 kommendes Eingangs-Audiosignal A1 von einem Wandler 2, welcher einen Verstärker 21, einen
15 Digital-Analog-Wandler 22 und einen Lautsprecher 23 aufweist, verstärkt, in ein analoges elektrisches Signal gewandelt und an dem Lautsprecher 23 ausgegeben. Ein von einem nahen Ende (nahen Sprecher) kommendes Audiosignal A0, das sowohl Sprache als auch Rauschen enthalten kann, wird an einem Rückwandler 3, der ein Mikrofon 31, einen Analog-Digital-Wandler 32 und einen Verstärker 33 aufweist,
20 gegebenenfalls zusammen mit einem Echosignal A1' des Eingangs-Audiosignals A1, in ein digitales elektrisches Signal gewandelt und verstärkt, bevor es nach einer entsprechenden Nachverarbeitung zur Echo unterdrückung in der Echo reduktionseinheit 4 an dem Ausgangskanal 5 als Ausgangs-Audiosignal A2 zur Übertragung an das ferne Ende (den fernen Sprecher) ausgegeben wird.

25

Die erfindungsgemäße vorgesehene Echo reduktionseinheit 4 weist ein adaptives FIR-Filter (Finite-Impulsantwort-Filter) 40, das auch als Echokompensationseinheit bezeichnet werden kann, auf, das eine Schätzung e des in dem Ausgangssignal z des Rückwandlers 3 enthaltenen Echoes des Eingangs-Ausgangssignals A1 erzeugt. Dieses
30 Schätzsignal e wird zum einen von dem Ausgangssignal z des Rückwandlers 3 mittels

eines Subtrahierers 41 subtrahiert. Dabei können jedoch restliche Echos in dem Ausgangssignal r des Subtrahierers 41 verbleiben, einerseits wegen einer Fehljustierung des FIR-Filters 40 und andererseits, da das FIR-Filter 40 nur eine finite Echodauer von etwa 25 Millisekunden abschätzen kann.

5

Das echo reduzierte Ausgangssignal r des Subtrahierers 41 wird nachfolgend einer Nachverarbeitungseinheit 42, die auch als Echounterdrückungseinheit bezeichnet werden kann, zugeführt, die zunächst das Signal r mittels einer FFT-Einheit 421, insbesondere mit 128 Punkten, in den Frequenzraum transformiert. Restliche Echos

10 werden dann unterdrückt, indem Frequenzen mit dominanten Echobeträgen gedämpft werden. Frequenzen, bei denen das von dem nahen Ende kommende Audiosignal A_0 dominiert, werden dagegen weniger gedämpft. Diese Dämpfung erfolgt in dem Multiplizierer 422 mittels einer frequenzabhängigen Dämpfungsfunktion D , die mittels einer Dämpfungsberechnungseinheit 423, die vorzugsweise das Verfahren des mittleren
15 Fehlerquadrats (Minimum Mean Square Error MMSE) verwendet. Dazu ist ein frequenzabhängiges Echoschätzsignal E erforderlich, das aus dem zeitabhängigen Echoschätzsignal e durch FFT-Transformation mittels einer zu der FFT-Transformationseinheit 421 gleichwirkenden FFT-Transformationseinheit 424 berechnet wird. Da das zeitabhängige Echoschätzsignal e nur die frühen Echos,
20 beispielsweise innerhalb der ersten 25 Millisekunden, enthält, ist außerdem eine Echorefexionseinheit 425 (echo reverberation) vorgesehen, die eine Schätzung der späteren Echos unter Annahme eines exponentiellen Abfalls der Echoenergie hinzufügt.

25 Die Stärke der Echounterdrückung wird mittels eines Echounterdrückungsfaktors s gesteuert, mit dem das frequenzabhängige Echoschätzsignal E in dem Multiplizierer 426 multipliziert wird. Eine Sprachaktivitätsdetektionseinheit 427, die aus dem in den Frequenzbereich transformierten Ausgangssignal r des Subtrahierers 41 und dem in den Frequenzbereich transformierten Echoschätzsignal e detektiert, ob in dem von den nahen Ende kommenden Audiosignal A_0 ein Sprachsignal enthalten ist, gibt eine
30 entsprechende Information an eine Steuereinheit 6 weiter. Außerdem wird der

Steuereinheit 6 eine von einer weiteren Sprachaktivitätsdetektionseinheit 7 ermittelte Information zugeführt, ob in dem Eingangs-Audiosignal A1 ein Sprachsignal enthalten ist. Aus diesen Informationen, und optional einer weiteren Zeitinformation aus einer Zeitsteuereinheit 8, wird dann nach dem erfindungsgemäßen Verfahren der je nach

- 5 Situation geeignete Echounterdrückungsfaktor s berechnet, wie nachfolgend erläutert werden soll. Nachdem das in den Frequenzbereich transformierte Ausgangssignal r des Subtrahierers 41 mit der entsprechend berechneten Dämpfung D multipliziert wurde, wird es wieder in den Zeitbereich mittels einer inversen FFT-Transformationseinheit 428 transformiert, wonach es dann an das ferne Ende über den Ausgangskanal 5 als
- 10 Ausgangs-Audiosignal A2 übertragen wird.

Die Erfindung nutzt den psycho-akustischen Effekt aus, dass ein menschlicher Zuhörer stationäre Signale nicht als Echosignale empfindet. Eine einfache Erklärung dafür ist, dass das „Echosignal“ von Hintergrundrauschen, welches vom fernen Ende her kommt,

- 15 wie das Rauschen selbst klingt, da die spektralen Anteile von Rauschen sich normalerweise bei der Rückkopplung von dem Lautsprecher 23 zu dem Mikrofon 31 nicht verändern. Dadurch wird ein derartiges „Echo“ von dem Hintergrundrauschen maskiert und ist somit nicht hörbar. Außerdem klingt „Echo“ von Hintergrundrauschen, das von dem fernen Ende her kommt, wie Rauschen, das von irgendwo anders her
- 20 kommt, z. B. von dem nahen Ende. Ein Hörer am fernen Ende dürfte deshalb ein solches „Echo“ nicht wirklich als Echo ansehen, auch wenn er es hören könnte. Die Echounterdrückung kann deshalb in Zeitpunkten, in denen nur Hintergrundrauschen in dem von dem fernen Ende kommenden Eingangs-Audiosignal A1 enthalten ist, stark reduziert werden, ohne dass Echos von dem am fernen Ende befindlichen Teilnehmer
- 25 wahrgenommen werden.

Erfindungsgemäß reduziert deshalb die Steuereinheit 6 den Echounterdrückungsfaktor s auf einen kleinen Wert s_{low} in solchen Situationen. Diese Reduzierung wird erfindungsgemäß jedoch nicht abrupt, sondern schrittweise und kontinuierlich

- 30 vorgenommen, insbesondere entsprechend dem Abfall der Echoenergie. Ansonsten würde ein Abfallen des Echos hörbar sein, wenn der am fernen Ende befindliche

Sprecher aufhört zu sprechen. Die Steuereinheit 6 verwertet dabei, wie oben bereits angedeutet, die von den Sprachaktivitätsdetektionseinheiten 427 und 7 erhaltenen Informationen darüber, ob das der Nachverarbeitungseinheit 42 zugeführte Signal r bzw. das Eingangs-Audiosignal A1 ein Sprachsignal enthält. Das Verhalten, abhängig von dieser Information, ist in der nachfolgenden Tabelle zusammengestellt, wobei k die Rahmenzeit (frame time) ist und α den exponentiell verlaufenden Abfall der Echoenergie wiedergibt.

5 von dieser Information, ist in der nachfolgenden Tabelle zusammengestellt, wobei k die Rahmenzeit (frame time) ist und α den exponentiell verlaufenden Abfall der Echoenergie wiedergibt.

A1 = Sprachsignal	A0 = Sprachsignal	Echounterdrückungsfaktor s
Nein	-	$s[k] = \alpha \cdot s[k-1] + (1 - \alpha) \cdot s_{low}$
Ja	Ja	$s[k] = s_{double}$
Ja	Nein	$s[k] = s_{high}$

Typischerweise wird dieselbe Echowiedergabekonstante α auch in der Echowiedergabeeinheit 425 zur Echoabschätzung verwendet. Bevorzugt weist der Faktor α einen Wert im Bereich von 0,5 und 0,99, vorzugsweise im Bereich von 0,75 bis 0,85, auf. Bei einer typischen Rahmenzeit k von beispielsweise 0,01 Sekunden ist ein Wert von $\alpha=0,8$ vorteilhaft. Allgemein kann der Faktor α auch beschrieben werden mit $\alpha = 10^{(-3 \cdot \text{Rahmenzeit} / T_{60})}$, wobei T_{60} die Nachhallzeit eines typischen Raumes (z.B. eines Büros, Wohnzimmers, Autos) ist.

10 Echowiedergabeeinheit 425 zur Echoabschätzung verwendet. Bevorzugt weist der Faktor α einen Wert im Bereich von 0,5 und 0,99, vorzugsweise im Bereich von 0,75 bis 0,85, auf. Bei einer typischen Rahmenzeit k von beispielsweise 0,01 Sekunden ist ein Wert von $\alpha=0,8$ vorteilhaft. Allgemein kann der Faktor α auch beschrieben werden mit $\alpha = 10^{(-3 \cdot \text{Rahmenzeit} / T_{60})}$, wobei T_{60} die Nachhallzeit eines typischen Raumes (z.B. eines Büros, Wohnzimmers, Autos) ist.

15 Echowiedergabeeinheit 425 zur Echoabschätzung verwendet. Bevorzugt weist der Faktor α einen Wert im Bereich von 0,5 und 0,99, vorzugsweise im Bereich von 0,75 bis 0,85, auf. Bei einer typischen Rahmenzeit k von beispielsweise 0,01 Sekunden ist ein Wert von $\alpha=0,8$ vorteilhaft. Allgemein kann der Faktor α auch beschrieben werden mit $\alpha = 10^{(-3 \cdot \text{Rahmenzeit} / T_{60})}$, wobei T_{60} die Nachhallzeit eines typischen Raumes (z.B. eines Büros, Wohnzimmers, Autos) ist.

Durch die erfindungsgemäße Steuerung wird eine deutlich verbesserte Sprachqualität der von dem nahen Ende an das ferne Ende übertragenen Sprache erreicht, wobei für die Bestimmung des Echounterdrückungsfaktors bei fehlendem Sprachsignal in dem Eingangs-Audiosignal das Vorliegen eines Sprachsignals in dem von dem nahen Ende kommenden Audiosignal nicht von Bedeutung ist. Da das Vorliegen eines Sprachsignals in dem Eingangs-Audiosignal mit hoher Zuverlässigkeit mittels der Sprachaktivitätsdetektionseinheit 7 bestimmt werden kann, können verhältnismäßig kleine Werte für den Echounterdrückungsfaktor bei fehlendem Sprachsignal in dem Eingangs-Audiosignal eingestellt werden, die beispielsweise 0,5 oder weniger betragen, ohne das Risiko eingehen zu müssen, dass Echos fälschlicherweise nicht unterdrückt werden, was ebenfalls zu einer hohen Übertragungsqualität eines von dem nahen Ende

20 Eingangs-Audiosignal das Vorliegen eines Sprachsignals in dem von dem nahen Ende kommenden Audiosignal nicht von Bedeutung ist. Da das Vorliegen eines Sprachsignals in dem Eingangs-Audiosignal mit hoher Zuverlässigkeit mittels der Sprachaktivitätsdetektionseinheit 7 bestimmt werden kann, können verhältnismäßig kleine Werte für den Echounterdrückungsfaktor bei fehlendem Sprachsignal in dem Eingangs-Audiosignal eingestellt werden, die beispielsweise 0,5 oder weniger betragen, ohne das Risiko eingehen zu müssen, dass Echos fälschlicherweise nicht unterdrückt werden, was ebenfalls zu einer hohen Übertragungsqualität eines von dem nahen Ende

25 Eingangs-Audiosignal das Vorliegen eines Sprachsignals in dem von dem nahen Ende kommenden Audiosignal nicht von Bedeutung ist. Da das Vorliegen eines Sprachsignals in dem Eingangs-Audiosignal mit hoher Zuverlässigkeit mittels der Sprachaktivitätsdetektionseinheit 7 bestimmt werden kann, können verhältnismäßig kleine Werte für den Echounterdrückungsfaktor bei fehlendem Sprachsignal in dem Eingangs-Audiosignal eingestellt werden, die beispielsweise 0,5 oder weniger betragen, ohne das Risiko eingehen zu müssen, dass Echos fälschlicherweise nicht unterdrückt werden, was ebenfalls zu einer hohen Übertragungsqualität eines von dem nahen Ende

kommenden Sprachsignals beträgt. Auch wenn beide Sprecher gleichzeitig sprechen, kann die Übertragung des Sprachsignals vom nahen Sprecher verbessert werden, da die Sprachaktivitätsdetektionseinheit 7 das Umschalten auf den niedrigen Echounterdrückungsfaktor s_{low} während kurzer Pausen, sogar innerhalb eines

5 gesprochenen Wortes, der von dem fernen Ende kommenden Sprache erfolgen kann.

Das Zurückschalten auf den niedrigen Echounterdrückungswert s_{low} könnte auch um eine Totzeit verzögert werden, was aber während der gesamten Totzeit, die beispielsweise 0,4 Sekunden betragen könnte, unnötig hohe, weil gleichbleibende

10 Unterdrückung bewirkt. Besser ist deshalb das oben beschriebene exponentielle Abklingen von dem hohen Echounterdrückungswert s_{high} auf den niedrigen Echounterdrückungswert s_{low} . Dies bildet im wesentlichen den wahren Zeitverlauf der Echoenergie nach, da diese auch exponentiell mit der Zeit abnimmt.

15 Bei Verwendung eines adaptiven FIR-Filters 40, wie es vorliegend bevorzugt eingesetzt wird, wurde festgestellt, dass insbesondere beim Start der Kommunikationsübertragung, also beim Beginn eines Telefongesprächs, das FIR-Filter 40 noch nicht adaptiert ist und deshalb keine Echoabnahme vom Signal r nach z erzielt wird. Die Echoabnahme des FIR-Filters steigt erst langsam an, so dass zu Beginn noch Echos durchkommen, wenn

20 die Nachverarbeitungsvorrichtung 42 noch in normaler Art und Weise unterdrückt. Dies kann noch dadurch verschärft werden, dass die Nachverarbeitungsvorrichtung 42 Echos proportional zur Leistung des Echoschätzsignals e unterdrückt. Wenn das FIR-Filter 40 also wie üblich mit Null-Koeffizienten initialisiert wird, ist diese Leistung zunächst 0 und erreicht frühestens nach einer Wartezeit von etwa einer halben Sekunde den wahren

25 Wert der Echoleistung.

Um dieses Problem auszugleichen wird erfahrungsgemäß deshalb für ausreichend Leistung des Echoschätzsignals e während der Startphase gesorgt, indem die Koeffizienten des FIR-Filters 40 ungleich Null initialisiert werden. Beispielsweise kann

30 ein Koeffizient gleich 1 gesetzt werden. Bei der Echoabnahmeeinheit 4 der vorliegend verwendeten Art ist der Wertebereich der Koeffizienten auf -1 bis +1 beschränkt, so

dass mehrere Koeffizienten vorbesetzt werden müssen, um eine realistische Leistung im Echoschätzsignal e beim Start zu erreichen. Dabei ist zu beachten, dass die Koeffizienten ein breitbandiges System repräsentieren, damit die Nachverarbeitungseinheit 42 auch Echos bei allen Frequenzen unterdrücken kann.

5 Geeignet ist beispielsweise ein Vorbesetzen von acht aufeinanderfolgenden Koeffizienten des FIR-Filters mit den Werten $\{1, 1, 1, 1, -1, -1, 1, -1\}$. Somit erhält man beim Start eine 8-fache Leistung des Echoschätzsignals e im Vergleich zum Vorbesetzen eines einzelnen Koeffizienten. Zudem überträgt ein so vorbesetztes FIR-Filter alle relevanten Frequenzen etwa in gleicher Weise.

10 Bei einer bevorzugten Ausgestaltung der erfindungsgemäßen Vorrichtung erfolgt deshalb zusätzlich eine Echounterdrückung durch die Nachverarbeitungsvorrichtung 42 beim Start. Die Idee ist dabei, den hohen Echounterdrückungswert s_{high} entsprechend dem anfänglich schlechten FIR-Adaptionszustand zu erhöhen. Dazu kann beispielsweise die gezeigte Zeitsteuereinrichtung 8 verwendet werden. Der Wert des niedrigen Echounterdrückungswertes, also von s_{low} , wird beibehalten, da ohne ferne Sprache auch beim Start keine Echos entstehen.

15

Fig. 2 zeigt den Verlauf des Echounterdrückungsfaktors s , wobei der Verlauf des hohen Echounterdrückungswertes s_{high} gestrichelt dargestellt ist. Im Bereich A ist die Unterdrückung groß genug, um die zu geringe Echokompensation des FIR-Filters auszugleichen. In den Bereichen B und C findet keine FIR-Adaption wegen fehlender Sprache im Eingangs-Audiosignal statt, was durch das Ausgangssignal FE-SAD der Sprachaktivitätsdetektionseinheit 7 angezeigt wird. Deshalb wird der hohe Echounterdrückungswert s_{high} in den Bereichen B und C nicht weiter verkleinert; die Zeitsteuereinheit 8, die beispielsweise als Anfangszeitzähl器 ausgestaltet sein kann, zählt somit nicht weiter. Außerdem ist die Zeitsteuereinheit 8 bevorzugt so eingestellt, dass sich noch ein Sicherheitsbereich D ergibt, da die Adaptionsdauer abhängig von der fernen Sprache schwanken kann. Ab dem Bereich E ist der stationäre Zustand erreicht, und $s_{high} = s_{high,stat}$ bleibt konstant auf dem stationären hohen Echounterdrückungswert.

Erkennbar ist in Fig. 2 auch der in den Bereichen B und C, also bei fehlender Sprache im Eingangs-Audiosignal, erfindungsgemäß vorgesehene exponentielle Abfall des Echounterdrückungsfaktors s auf den niedrigen Echounterdrückungswert s_{low} . Des

- 5 weiteren ist in Fig. 2 in dem untersten Diagramm der Verlauf der Echo reduktion ER über der Zeit erkennbar. Wie ersichtlich ist nach etwa 0,8 Sekunden Adoptionszeit die erforderliche Genauigkeit sicher erreicht (bei einem vorliegend verwendeten FIR-Filter mit 200 Koeffizienten). Bei gegebener FIR-Filterlänge ist also die Konvergenzzeit bekannt. Mit $s < 4$ reagiert die Nachverarbeitungseinheit 42 auch auf geringe Echos mit
- 10 Maximalunterdrückung. In diesem Fall verhält sie sich wie eine frequenzabhängige Halbduplex-Steuerung in Senderichtung, d. h. es kommen keine Echos durch, aber während ferner Sprache ist auch die Senderichtung gesperrt. Mit $s_{high,init} = 10$ ist dies in den Bereichen A der Fall. In Bereich B wird die Unterdrückung entsprechend dem Nachhall exponentiell vermindert, und im Bereich C ist mit $s_{low} = 0,5$ die Senderichtung
- 15 freigeschaltet. Ab dem Bereich E wird mit $s_{high,stat} = 1,5$ nur noch die Unterdrückung von Restechos, die nicht vom FIR-Filter kompensiert wurden, erreicht, dafür aber die Senderichtung nie ganz gesperrt, d. h. das Normalverhalten nach der Initialisierungsphase ist erreicht.
- 20 Ein Blockschaltbild einer beispielhaften Sprachaktivitätsdetektionseinheit, wie sie erfindungsgemäß verwendet wird, ist in Figur 3 gezeigt. Eine derartige Einheit kann ein Sprachsignal innerhalb eines Signal mit stationärem Hintergrundrauschen detektieren. Diese weist in der gezeigten Ausführungsform eine Einheit 71 zur Bestimmung des über einen kurzen Zeitraum gemessenen Signalpegels L_s , eine Einheit 72 zur Bestimmung
- 25 des über einen längeren Zeitraum gemessenen minimalen Signalpegels M_1 des Eingangssignals I sowie einen Komparator 73 auf. Dabei wird ausgenutzt, dass ein Sprachsignal einen stark zeitabhängigen Signalpegel im Vergleich zu einem Hintergrundrauschsignal aufweist. Immer wenn der über einen kurzen Zeitraum gemessene Signalpegel L_s um einen Faktor C, beispielsweise $C = 2$ (was 6 dB entspricht) größer ist als der minimale über einen längeren Zeitraum (z. B. 2 Sekunden) gemessene Signalpegel M_1 ist, wird Sprachaktivität detektiert und ein entsprechendes

Ausgangssignal O mit dieser Information ausgegeben.

Sowohl das vorzugsweise exponentielle Abklingen von s_{high} nach s_{low} , als auch die Steuerung von s_{high} beim Start erlauben es, die Unterdrückung so gering wie möglich zu 5 halten, aber dennoch alle Echos zu entfernen. Je geringer die Unterdrückung ist, desto besser ist die Vollduplex-Kommunikation bei Gegensprechen (double-talk). Vollduplex-Kommunikation heißt hier, dass der ferne Sprecher hört, wenn ihn der nahe Sprecher unterbrechen will oder eine kurze Bemerkung macht, während der ferne Sprecher aktiv ist. Ein weiterer Vorteil der erfindungsgemäßen Steuerung von s_{high} beim 10 Start ist, dass am Anfang eines Telefongesprächs Echos mit absoluter Sicherheit vermieden werden. Während dieser Zeit ist die Vollduplex-Fähigkeit vom nahen zum fernen Sprecher hin zwar eingeschränkt, was aber kein Nachteil ist, da in den ersten Sekunden eines Telefonats selten Gegensprechen stattfindet, sondern jeder Teilnehmer zunächst seinen Namen oder einen Gruß ausspricht.

PATENTANSPRÜCHE

1. Vorrichtung zur Reduktion von Echo für eine Anordnung zum Übertragen von Audiosignalen, insbesondere von gesprochener Sprache, mit
 - einer Echoreduktionseinheit (4), die zwischen einem Eingangskanal (1) zum Empfangen eines von einem fernen Ende kommenden Eingangs-Audiosignals (A1) und einem Ausgangskanal (5) zum Ausgeben eines Ausgangs-Audiosignals (A2) angeordnet ist, zur Unterdrückung eines in dem Ausgangs-Audiosignal (A2) enthaltenen Echosignals,
 - einer Sprachaktivitätsdetektionseinheit (7) zur Detektion eines in dem Eingangs-Audiosignal (A1) enthaltenen Sprachsignals, und
 - 10 - einer Steuereinheit (6) zur Einstellung eines Echounterdrückungsfaktors (s) der Echoreduktionseinheit (4) für die Unterdrückung von Echo,
dadurch gekennzeichnet,
dass die Steuereinheit (6) derart ausgebildet ist, dass der Echounterdrückungsfaktor (s) von einem während des Vorliegens eines Sprachsignals in dem Eingangs-Audiosignal 15 (A1) eingestellten hohen Echounterdrückungswert schrittweise und kontinuierlich auf einen niedrigen Echounterdrückungswert reduziert wird, wenn von der Sprachaktivitätsdetektionseinheit (7) detektiert wird, dass das Eingangs-Audiosignal (A1) kein Sprachsignal enthält.
- 20 2. Vorrichtung nach Anspruch 1,
dadurch gekennzeichnet,
dass die Steuereinheit (6) derart ausgebildet ist, dass die Reduzierung des Echounterdrückungsfaktors (s) von dem hohen Echounterdrückungswert auf den niedrigen Echounterdrückungswert exponentiell erfolgt.

3. Vorrichtung nach Anspruch 2,
dadurch gekennzeichnet,
dass die Reduzierung des Echounterdrückungsfaktors (s) entsprechend der Funktion
5 $s[k] = \alpha \cdot s[k - 1] + (1 - \alpha) \cdot s_{low}$ erfolgt, wobei $s[k]$ der Echounterdrückungswert zum
Zeitpunkt k, α ein das exponentielle Reduzierungsverhalten repräsentierender Faktor
und s_{low} der minimale Echo-Unterdrückungswert sind.

4. Vorrichtung nach Anspruch 3,
10 dadurch gekennzeichnet,
dass der minimale Unterdrückungswert s_{low} einen Wert im Bereich von 0,1 bis 1,
vorzugsweise von etwa 0,5, und der die exponentielle Reduzierung repräsentierende
Faktor α einen Wert im Bereich von 0,5 und 0,99, vorzugsweise im Bereich von 0,75
bis 0,85, aufweist.

15 5. Vorrichtung nach Anspruch 1,
dadurch gekennzeichnet,
dass die Steuereinheit (6) ausgebildet ist zur zeitlichen Verzögerung der Reduzierung
des Echounterdrückungsfaktors (s) von dem eingestellten hohen Echounterdrückungs-
20 wert auf den niedrigen Echounterdrückungswert um eine Totzeit, insbesondere im
Bereich von 0,1 bis 1 Sekunde, vorzugsweise von etwa 0,4 Sekunden.

6. Vorrichtung nach Anspruch 1,
dadurch gekennzeichnet,
25 dass eine zweite Sprachaktivitätsdetektionseinheit (427) zur Detektion eines in dem der
Echoreduktionseinheit (4) zugeführten echoreduzierten Audiosignal (r) enthaltenen, von
einem nahen Ende kommenden Sprachsignals vorgesehen ist und dass die Steuereinheit
(6) derart ausgebildet ist, dass der Echounterdrückungsfaktor (s) auf den hohen
Echounterdrückungswert eingestellt wird, wenn das Eingangs-Audiosignal ein
30 Sprachsignal und das echoreduzierte Audiosignal (r) kein von dem nahen Ende

5 kommendes Sprachsignal enthalten, und dass der Echounterdrückungsfaktor (s) auf einen mittleren, zwischen dem hohen und dem niedrigen Echounterdrückungswert liegenden Echounterdrückungswert eingestellt wird, wenn das Eingangs-Audiosignal (A1) ein Sprachsignal und das echo reduzierte Audiosignal (r) ein von dem nahen Ende

5 kommendes Sprachsignal enthalten.

7. Vorrichtung nach Anspruch 1,
dadurch gekennzeichnet,
dass die Echoreduktionseinheit (4) ein adaptives FIR-Echofilter (40) aufweist zur
10 Bestimmung eines Echoschätzsignals (e) und dass die Koeffizienten des adaptiven FIR-
Echofilters (40) beim Start des Empfangs des Eingangs-Audiosignals (A1) auf einen
Wert ungleich von Null initialisiert werden.

8. Vorrichtung nach Anspruch 7,
15 dadurch gekennzeichnet,
dass die Steuereinheit derart (6) ausgebildet ist, dass der Echounterdrückungsfaktor (s)
beim Start des Empfangs des Eingangs-Audiosignals (A1) auf einen maximalen Start-
Echounterdrückungswert eingestellt wird und danach kontinuierlich, insbesondere
linear, während des Vorhandenseins eines Sprachsignals in dem Eingangs-Audio signal
20 (A1) auf einen stationären hohen Echo-Unterdrückungswert reduziert wird.

9. Vorrichtung nach Anspruch 8,
dadurch gekennzeichnet,
dass die Zeitdauer für die Reduzierung des Echounterdrückungsfaktors (s) von dem
25 maximalen Start-Echounterdrückungswert auf den statischen hohen Echounter-
drückungswert so eingestellt ist, dass sie in etwa der Adaptionsdauer des adaptiven FIR-
Echofilters (40) entspricht.

10. Vorrichtung nach Anspruch 9,

dadurch gekennzeichnet,

dass der maximale Start-Echounterdrückungswert im Bereich zwischen 5 und 15, insbesondere um etwa 10, liegt und dass der statische hohe Echounterdrückungswert

5 etwa im Bereich von 1 bis 2, insbesondere etwa im Bereich von 1,5, liegt.

11. Verfahren zur Reduktion von Echo bei der Übertragung von Audiosignalen,

insbesondere gesprochener Sprache, mit den Schritten:

- Echounterdrückung zur Unterdrückung eines in einem Ausgangs-Audiosignal (A2) 10 enthaltenen Echosignals (A1) eines Eingangs-Audiosignals (A1),
- Detektion eines in dem Eingangs-Audiosignal (A1) enthaltenen Sprachsignals, und
- Einstellung eines Echounterdrückungsfaktors (s) für die Unterdrückung von Echo, dadurch gekennzeichnet,

dass der Echounterdrückungsfaktor (s) von einem während des Vorliegens eines 15 Sprachsignals in dem Eingangs-Audiosignal (A1) eingestellten hohen Echounterdrückungswert schrittweise und kontinuierlich auf einen niedrigen Echounterdrückungswert reduziert wird, wenn detektiert wird, dass das Eingangs-Audiosignal (A1) kein Sprachsignal enthält.

20 12. Computerprogramm mit Computerprogrammmitteln zur Veranlassung eines Computers zur Ausführung der Verfahrensschritte des Verfahrens nach Anspruch 11, wenn das Verfahren auf einem Computer ausgeführt wird.

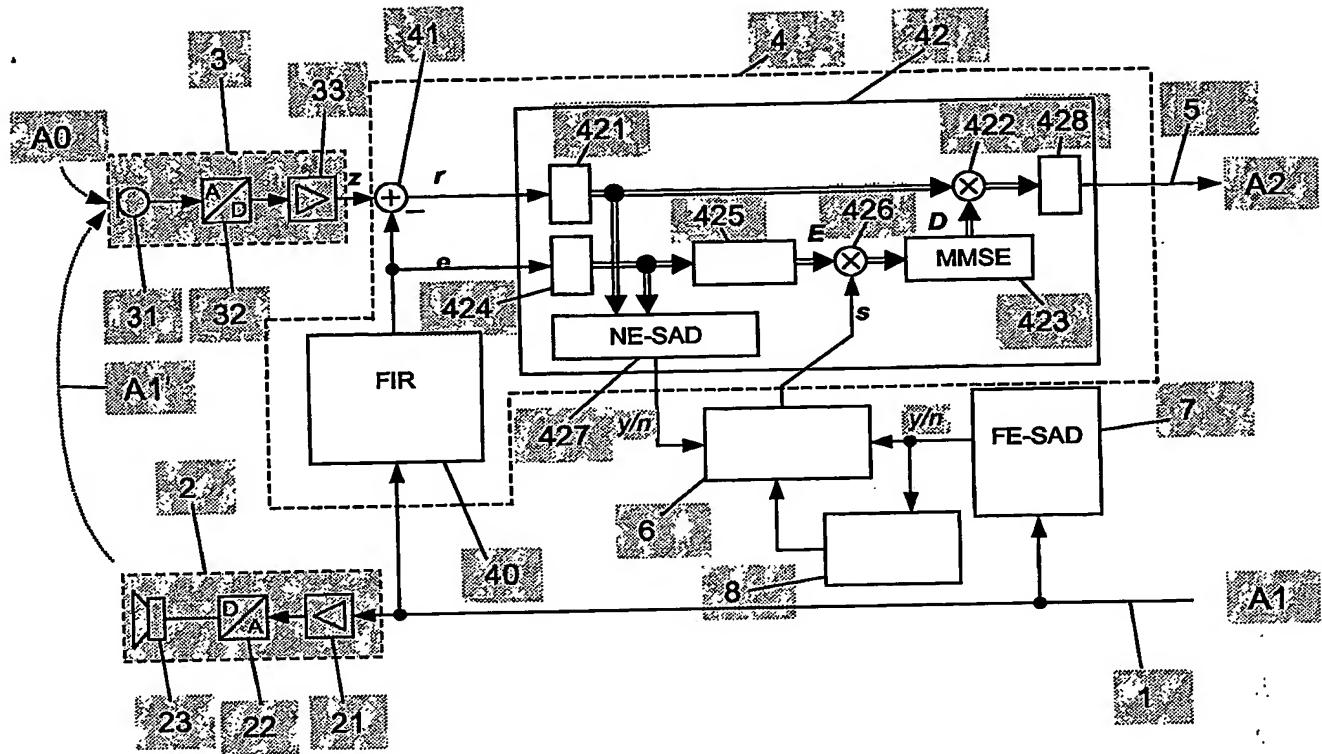


Fig. 1

2/2

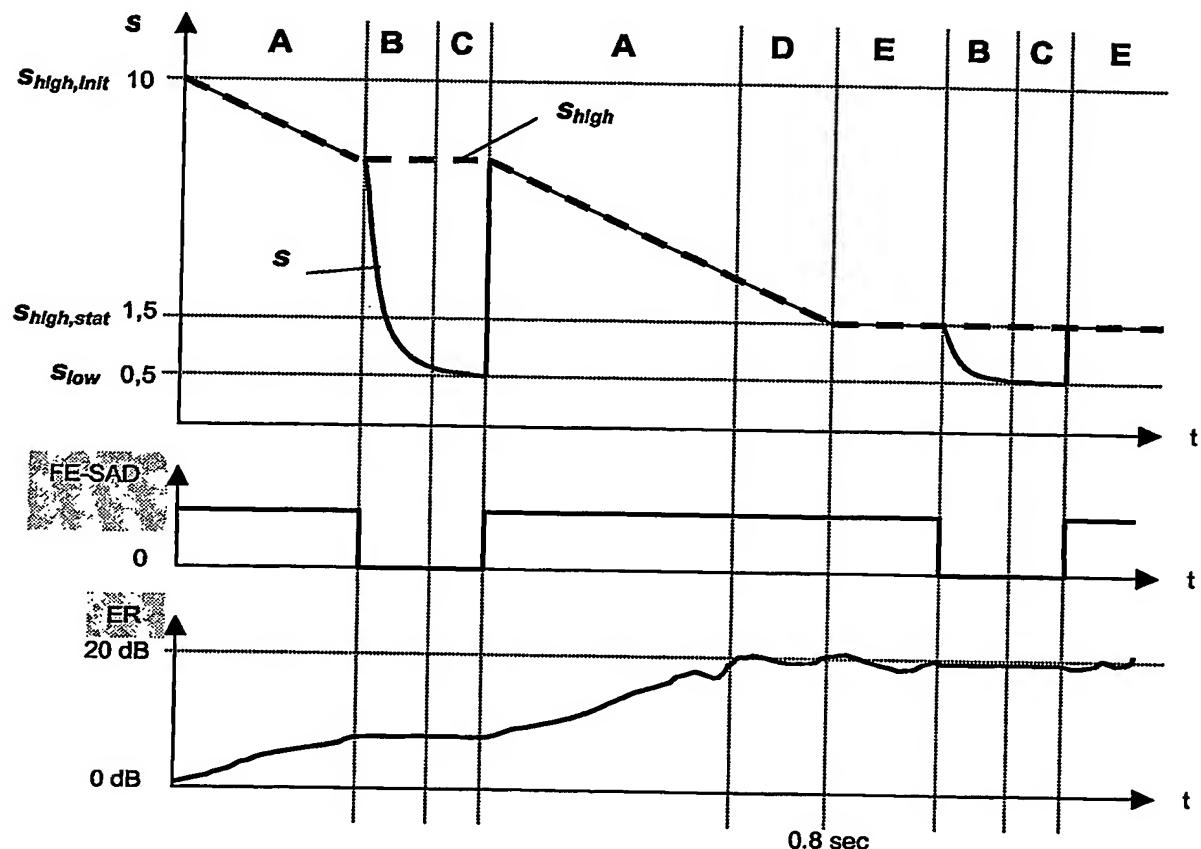


Fig. 2

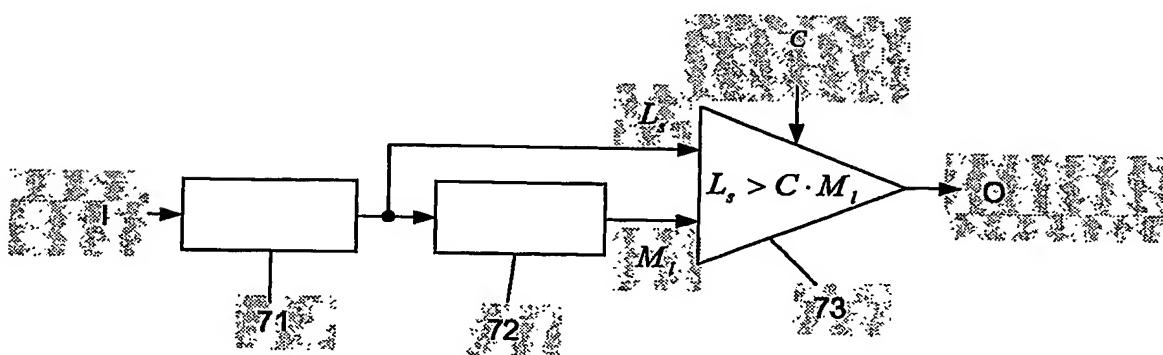


Fig. 3

THIS PAGE BLANK (USPTO)

PCT Application
IB0306252

